

Genes for Tryptophan Biosynthesis in the Halophilic Archaeobacterium *Haloferax volcanii*: the *trpDFEG* Cluster

WAN L. LAM, SUSAN M. LOGAN, AND W. FORD DOOLITTLE*

Canadian Institute for Advanced Research and Department of Biochemistry, Dalhousie University,
Halifax, Nova Scotia, Canada B3H 4H7

Received 17 July 1991/Accepted 20 December 1991

Tryptophan auxotrophs of the archaeobacterium *Haloferax volcanii* define a cluster of overlapping genes homologous to eubacterial-eukaryotic *trpD*, *-F*, *-E*, and *-G*, linked in that order and each preceded by a possible ribosome binding site. Residues involved in feedback inhibition of eubacterial anthranilate synthetases are conserved.

The organization of genes for tryptophan biosynthesis has been studied in many eubacteria and eukaryotes (5, 6). Comparable information from archaeobacteria remains limited. Meile et al. (18) recently completed the sequence of the *Methanobacterium thermoautotrophicum* gene cluster, finding open reading frames corresponding to all activities in the order *trpEGCFBAD*. We used an ordered set of cosmids bearing *Haloferax volcanii* DNA to transform tryptophan auxotrophs of this halophilic archaeobacterium to prototrophy and localized all of 29 mutations to one of two regions of the genome (15). Sequencing of the first region (covering 10 mutations) revealed the overlapping genes *trpCBA*.

The 19 remaining mutations could be transformed to prototrophy with cosmid G203, 1,200 kbp away from the *trpCBA* cluster on the *H. volcanii* chromosome (2). Each mutation could also be transformed with p trpE12 , obtained by shotgun cloning wild-type DNA into our shuttle vector pWL102 (16). A 4-kbp *Hind*III subfragment was cloned into M13 vectors and sequenced (13, 19).

The region complementing anthranilate and indole-utilizing tryptophan auxotrophs contains four open reading frames (Fig. 1), each with significant sequence similarity to one of the tryptophan enzymatic domains defined by Hutter et al. (14).

The first gene of the cluster, *trpD*, encodes phosphoribosyltransferase (337 amino acids). All but a few of the conserved positions identified by Crawford and Milkman in an alignment of many eubacterial TrpD sequences are found in the *H. volcanii* sequence (5, 6), which shows 34 and 24% amino acid identity with the TrpD proteins of *Escherichia coli* and *Saccharomyces cerevisiae*, respectively.

H. volcanii trpF should encode *N*-phosphoribosylanthranilate isomerase (22 amino acids). Residues of presumed functional importance (6, 21) are largely conserved, and this protein is about equally similar to *E. coli* and *S. cerevisiae* homologs (30 and 32% amino acid identity, respectively).

The *H. volcanii trpE* gene which encodes the α subunit of anthranilate synthase (523 amino acids), does not begin with AUG. Overlapping the stop codon of the upstream *trpF* is an alternate start signal (GUGA). Meile et al. (18) also infer a GUG start for *M. thermoautotrophicum trpG*, and this gene too is preceded by *trpE*, but the extent of the overlap and the phase of the frameshift differ. The conserved segment between positions 52 and 55 of *trpE* (Leu-Leu-Glu-Ser), iden-

tified as a site for mutations affecting feedback inhibition by tryptophan in *Brevibacterium lactofermentum* (17) and *Salmonella typhimurium* (1), is found in the *H. volcanii* sequence. The halobacterial TrpE shows 29 and 30% amino acid sequence identity with *E. coli* and yeast enzymes, respectively, 34% identity with its *Methanobacterium* homolog, and comparable values with the homologous *pabB* and *phnA* genes of *E. coli* and *Pseudomonas aeruginosa* (Fig. 2). *H. volcanii* provides the only example of *trpE* embedded within a gene cluster—in all other organisms, this gene, encoding the first activity in the biochemical pathway, is the first gene of its cluster or is solitary (5, 6).

TrpG, the β subunit of anthranilate synthase (204 amino acids), is the smallest of the seven *H. volcanii* Trp proteins. It is 30 and 37% identical in sequence to *E. coli* and yeast homologs, respectively, and 38% similar to *M. thermoautotrophicum trpG*.

The G+C content of the *trpDFEG* genes (between 67 and 74 mol%) is similar to that of bulk *Haloferax* chromosomal DNA. Codons ending with C or G account for more than 90% of the fourfold-degenerate third positions of codons for alanine, glycine, proline, threonine, and valine. C is much preferred to G in the third position, except for codons with a C in the second position, as observed for the *H. volcanii hisC* locus (3).

All seven Trp enzymes of *H. volcanii* contain 17 to 20% acidic residues (aspartic acid and glutamic acid). Eisenberg and coworkers have argued that acidic residues stabilize proteins in the cytoplasm of halophilic archaeobacteria, which is nearly saturated with KCl (7, 24).

The *trpD*, *-F*, *-E*, and *-G* genes overlap by their stop and start codons (as do the genes of the *trpCBA* cluster). All four genes in the *trpDFEG* cluster are preceded (5 to 7 bp upstream of the ATG) by good potential ribosome-binding sequences: GGUGAU for *trpD*, GGAGG for *trpF*, AGG AGGU for *trpE*, and AGGAGG for *trpG*. The sequence of the 3' terminus of the *H. volcanii* small subunit rRNA is 3'-UCCUCCACUA (12). About 40 bp in front of the translation start of *trpD* is the sequence ATTTGTA, which resembles the archaeobacterial box A promoter consensus [TTTA(A/T)A] (22, 25). A similar sequence (TTATGTA) is also found approximately 30 bp upstream from the start codon of *trpCBA*. No close approximations to these presumed promoter sequences are found within 100 bp upstream of any internal open reading frame in the *trpDFEG* cluster, but transcript analysis is required to exclude the

* Corresponding author.

TrpE-like sequences

PhnA	56
PhaB	56
PAEE	56
PAEG	56
ECOE	56
ECOG	56
MTHG	56
HVOE	56
SCEG	56
PhnA	89
PhaB	89
PAEE	89
PAEG	89
ECOE	89
ECOG	89
MTHG	89
HVOE	89
SCEG	89
PhnA	139
PhaB	139
PAEE	139
PAEG	139
ECOE	139
ECOG	139
MTHG	139
HVOE	139
SCEG	139
PhnA	197
PhaB	197
PAEE	197
PAEG	197
ECOE	197
ECOG	197
MTHG	197
HVOE	197
SCEG	197
PhnA	249
PhaB	249
PAEE	249
PAEG	249
ECOE	249
ECOG	249
MTHG	249
HVOE	249
SCEG	249
PhnA	306
PhaB	306
PAEE	306
PAEG	306
ECOE	306
ECOG	306
MTHG	306
HVOE	306
SCEG	306
PhnA	357
PhaB	357
PAEE	357
PAEG	357
ECOE	357
ECOG	357
MTHG	357
HVOE	357
SCEG	357

PhnA	417
PhaB	417
PAEE	417
PAEG	417
ECOE	417
ECOG	417
MTHG	417
HVOE	417
SCEG	417
PhnA	476
PhaB	476
PAEE	476
PAEG	476
ECOE	476
ECOG	476
MTHG	476
HVOE	476
SCEG	476
PhnA	523
PhaB	523
PAEE	523
PAEG	523
ECOE	523
ECOG	523
MTHG	523
HVOE	523
SCEG	523
PhnA	60
PhaB	60
PAEE	60
PAEG	60
ECOE	60
ECOG	60
MTHG	60
HVOE	60
SCEG	60
PhnA	120
PhaB	120
PAEE	120
PAEG	120
ECOE	120
ECOG	120
MTHG	120
HVOE	120
SCEG	120
PhnA	176
PhaB	176
PAEE	176
PAEG	176
ECOE	176
ECOG	176
MTHG	176
HVOE	176
SCEG	176
PhnA	204
PhaB	204
PAEE	204
PAEG	204
ECOE	204
ECOG	204
MTHG	204
HVOE	204
SCEG	204

FIG. 2. Alignment of *trpE*, *trpG*, and related sequences. Inferred translation products of *H. volcanii trpE* and *trpG* (HVOE and HVOG) genes were aligned with *trpE* and *trpG* gene products from *P. aeruginosa* (PAEE and PAEG) (9), *E. coli* (ECOE and ECOG) (23), *M. thermoautotrophicum* (MTHG and MTHG) (18), and *S. cerevisiae* (SCEE and SCEG) (14) with the MULTALIN software of Corpet (4). Also aligned are the *E. coli pabA* and *pabB* and *P. aeruginosa phnA* and *phnB* gene products (8, 10). These two sets of genes code for a subunit of *p*-aminobenzoate synthase and a component of the phenazine pathway, respectively, which are homologs of anthranilate synthase subunits (5, 6). Asterisks mark residues shared by all *trp* gene products shown.

independently in several eubacterial and archaeobacterial lineages to give rise to the patterns now observed. By either view, we may surely count the clustering of genes for related functions as one feature in which archaeobacteria resemble eubacteria.

Nucleotide sequence accession number. The sequences described in this report will appear in the EMBL GenBank with the accession number M83788.

This work was supported by grants from the Medical Research Council of Canada and the U.S. Office of Naval Research. W.F.D. is a Fellow of the Canadian Institute for Advanced Research.

We thank Keith Conover for help with computer analysis of sequence data and Arlin Stoltzfus for a careful reading of the manuscript.

REFERENCES

- Caliguri, M. G., and R. Bauerle. 1991. Identification of amino acid residues involved in feedback regulation of the anthranilate

synthase complex from *Salmonella typhimurium*. Evidence for amino terminal regulatory site. *J. Biol. Chem.* **266**:8328-8335.

- Charlebois, R. L., L. C. Schalkwyk, J. D. Hoffman, and W. F. Doolittle. 1991. Detailed physical map and set of overlapping clones covering the genome of the archaeobacterium *Haloferax volcanii* DS2. *J. Mol. Biol.* **222**:509-524.
- Conover, R. K., and W. F. Doolittle. 1990. Characterization of a gene involved in histidine biosynthesis in *Halobacterium (Haloferax) volcanii*: isolation and rapid mapping by transformation of an auxotroph with cosmid DNA. *J. Bacteriol.* **172**:3244-3249.
- Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**:10881-10890.
- Crawford, I. P. 1989. Evolution of a biosynthetic pathway: the tryptophan paradigm. *Annu. Rev. Microbiol.* **43**:567-600.
- Crawford, I. P., and R. Milkman. 1990. Orthologous and paralogous divergence, reticulate evolution, and lateral gene transfer in bacterial *trp* genes. In R. K. Selander, A. G. Clark, and T. S. Whittman (ed.), *Evolution at the molecular level*. Sinauer Associates, Inc., Sunderland, Mass.

7. Eisenberg, H., and E. J. Wachtel. 1987. Structural studies of halophilic proteins, ribosomes, and organelles of bacteria adapted to extreme salt concentrations. *Annu. Rev. Biophys. Biophys. Chem.* **16**:69–92.
8. Essar, D. W., L. Eberly, A. Hadero, and I. P. Crawford. 1990. Identification and characterization of genes for a second anthranilate synthase in *Pseudomonas aeruginosa*: interchangeability of the two anthranilate synthases and evolutionary implications. *J. Bacteriol.* **172**:884–900.
9. Essar, D. W., L. Eberly, C.-Y. Han, and I. P. Crawford. 1990. DNA sequences and characterization of four early genes of the tryptophan pathway in *Pseudomonas aeruginosa*. *J. Bacteriol.* **172**:853–866.
10. Goncharoff, P., and B. P. Nichols. 1984. Nucleotide sequence of *Escherichia coli* *pabB* indicates a common evolutionary origin of *p*-aminobenzoate synthetase and anthranilate synthetase. *J. Bacteriol.* **159**:57–62.
11. Gropp, F., P. Palm, and W. Zillig. 1989. Expression and regulation of *Halobacterium halobium* phage ϕ H genes. *Can. J. Microbiol.* **35**:182–188.
12. Gupta, R., J. M. Lanter, and C. R. Woese. 1983. Sequence of 16S ribosomal RNA from *Halobacterium volcanii*. *Science* **221**:656–659.
13. Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**:351–359.
14. Hutter, R., P. Niederberger, and J. A. DeMoss. 1986. Tryptophan biosynthetic genes in eukaryotic microorganisms. *Annu. Rev. Microbiol.* **40**:55–77.
15. Lam, W. L., A. Cohen, D. Tsoulhas, and W. F. Doolittle. 1990. Genes for tryptophan biosynthesis in the archaeobacterium *Haloferax volcanii*. *Proc. Natl. Acad. Sci. USA* **87**:6614–6618.
16. Lam, W. L., and W. F. Doolittle. 1989. Shuttle vectors for the archaeobacterium *Halobacterium volcanii*. *Proc. Natl. Acad. Sci. USA* **86**:5478–5482.
17. Matsui, K., K. Miwa, and K. Sano. 1987. Two single-base-pair substitutions causing desensitization to tryptophan feedback inhibition of anthranilate synthase and enhanced expression of tryptophan genes of *Brevibacterium lactofermentum*. *J. Bacteriol.* **169**:5330–5332.
18. Meile, L., R. Stettler, R. Banholzer, M. Kotik, and T. Leisinger. 1991. Tryptophan gene cluster of *Methanobacterium thermoautotrophicum* Marburg: molecular cloning and nucleotide sequence of a putative *trpEGCFBAD* operon. *J. Bacteriol.* **173**:5017–5023.
19. Messing, J. 1983. New M13 vectors for cloning. *Methods Enzymol.* **101**:20–78.
20. Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**:2444–2448.
21. Priestle, J. P., M. G. Grutter, J. L. White, M. G. Vincent, M. Kania, E. Wilson, T. S. Jardetzky, K. Kirschner, and J. N. Jansonius. 1987. Three-dimensional structure of the bifunctional enzyme N-(5'-phosphoribosyl)anthranilate isomerase-indole-3-glycerol-phosphate synthase from *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **84**:5690–5694.
22. Reiter, W. D., U. Hudepohl, and W. Zillig. 1990. Mutational analysis of an archaeobacterial promoter: essential role of a "TATA box" for transcription efficiency and start site selection *in vitro*. *Proc. Natl. Acad. Sci. USA* **87**:9509–9513.
23. Yanofsky, C., and I. P. Crawford. 1987. The tryptophan operon, p. 1453–1472. In F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
24. Zaccal, G., and H. Eisenberg. 1990. Halophilic proteins and the influence of solvent on protein stabilization. *Trends Biochem. Sci.* **15**:333–337.
25. Zillig, W., P. Palm, W.-D. Reiter, F. Gropp, G. Puhler, and H. Klenk. 1988. Comparative evaluation of gene expression in archaeobacteria. *Eur. J. Biochem.* **173**:473–482.